4.25 Applications of Molecular Dynamics Simulations in Drug Design

C Oostenbrink and M M H van Lipzig, Vrije Universiteit, Amsterdam, The Netherlands
W F van Gunsteren, Eidgenössische Technische Hochschule, Zürich, Switzerland

4.25.1 Biomolecular Simulation

Computer simulations on models of physical systems have been carried out for more than 50 years now. The first Monte Carlo simulations on liquids represented by spheres and hard disks were reported in 1953.[1] The classical equations of motion for such systems were first solved in a molecular dynamics simulation in 1957,[2] after which it still took several years to simulate a Lennard-Jones fluid in 1964[3] and liquid water in 1971.[4] Only 6 years later was the first molecular dynamics simulation of a protein published opening the way to true biomolecular simulation.[5] After this landmark, the field has seen tremendous improvements, both in terms of methods, algorithms, and parameters that describe the physics behind the models[6–9] and in terms of system sizes, system complexities, and simulation timescales.[10,11]

These developments have turned molecular dynamics simulations into a valuable tool, complementary to experimental investigations, to probe into the structure, dynamics, and activity of large biologically relevant molecules and molecular complexes. After careful validation of the methods and parameters that are involved, it offers the possibility to explore such biomolecular systems at a time and space resolution that is often inaccessible to experiment. It has been proven invaluable in the structure determination of proteins and nucleic acids from both x-ray and nuclear magnetic resonance (NMR) experiments and is regularly used to explore the flexibility and dynamical behavior of such molecules. The thermodynamic information that can be obtained from computer simulations allows for an analysis and understanding of it in terms of molecular processes and for a prediction of molecular properties. More and more often, molecular dynamics techniques are being used by medicinal chemists and become integrated in the multidisciplinary research of molecular medicine. Molecular flexibility and the rigorous inclusion of entropic terms in the computational estimate of binding affinity are more and more being recognized as essential parts of computational modeling.

Free energy can be seen as the driving force of virtually all molecular processes. Already at an early stage, simulations were used to calculate free energies and free enthalpies of molecular systems.[12-15] The underlying thermodynamic and statistical–mechanical theory had been developed years earlier[16,17] and many methods to calculate these thermodynamic quantities have been suggested over the years.[18-28] This chapter briefly reviews the numerical integration of the equations of motion and the statistical mechanics behind the simulations. It then continues to discuss several practical methods to calculate free energies from the simulations, focusing mainly on methods that have a theoretical foundation in statistical mechanics. To give a specific example, we will discuss various aspects of drug design where molecular dynamics simulations can make significant contributions.

4.25.2   Equations of Motion

Consider a system that contains N particles that are treated explicitly. Often, these particles are atoms, but also groups of atoms can be treated as a single particle, such as for instance an aliphatic CH, CH_2, or CH_3 group. In classical simulation, such a system is fully defined by the positions, r, and the conjugate momenta, p, of the individual particles, where r and p represent $3N$ dimensional vectors. We will use r_i and p_i for the three-dimensional vectors describing the position and momentum of particle i. In the absence of constraints and velocity dependent forces, the Hamiltonian of the system can be written as

$$H(\mathbf{r}, \mathbf{p}) = K(\mathbf{p}) + V(\mathbf{r}) \tag{1}$$

where $K(\mathbf{p})$ is the kinetic energy which can be calculated from

$$K(\mathbf{p}) = \sum_i \frac{\mathbf{p}_i^2}{2m_i} = \sum_i \tfrac{1}{2} m_i \mathbf{v}_i^2 \tag{2}$$

Here, we have used the definition of the momentum $\mathbf{p}_i \equiv m_i \mathbf{v}_i$, with m_i the mass of particle i and the velocity $\mathbf{v}_i$ the time derivative of the position of the particle, $\mathbf{v}_i = d\mathbf{r}_i / dt$.

$V(\mathbf{r})$ in eqn [1] is the potential energy, describing the interactions between the particles in the system and possibly external influences on the system. It is a function of the particle positions $\mathbf{r}$. The functional form and parameters describing $V(\mathbf{r})$ is called a force field. There are several well-known force fields for biomolecular simulation described in the literature, such as AMBER,[29-31] CHARMM,[32-34] CHARMm,[35] ECEPP/3,[36] ENCAD,[37,38] GROMOS,[39-41] and OPLS.[42,43] Assuming that the potential energy has been properly defined and parameterized, one can calculate the force f_i on particle i as the negative derivative of the potential energy with respect to its position,

$$\mathbf{f}_i = -\frac{\partial}{\partial \mathbf{r}_i} V(\mathbf{r}) \tag{3}$$

Using this force we can write down the equation of motion according to Newton,[44] which states that the second time derivative, or acceleration, of the position of a classical particle is equal to the force exerted on it divided by its mass,

$$\frac{d^2}{dt^2} \mathbf{r}_i = \frac{1}{m_i} \mathbf{f}_i \tag{4}$$

One can now integrate the equations of motion for all particles simultaneously and follow the movements of the particles in time. One, but not the only, way of doing so is through the leapfrog algorithm,[45] where the positions and velocities are propagated numerically at shifted times using a time step Δt,

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t + \tfrac{1}{2}\Delta t) \cdot \Delta t \tag{5}$$

$$\mathbf{v}_i(t + \tfrac{1}{2}\Delta t) = \mathbf{v}_i(t - \tfrac{1}{2}\Delta t) + m_i^{-1} \mathbf{f}_i(t) \cdot \Delta t \tag{6}$$

The application of eqns [5] and [6] will lead to a trajectory of all particle positions in the system. Alternative integration schemes such as the Verlet[46] or Beeman[47] algorithms can be shown to produce the same positional trajectory.[48] The time step should be chosen sufficiently small to correctly integrate the fastest motion in the system. In biomolecular simulation, the fastest motions would typically be the bond vibrations, requiring a time step of 0.5 fs. For this reason, bonds are often not treated as flexible degrees of freedom, but the bond lengths are constrained to a given value.[49] Bond length constraints do not greatly affect the overall dynamics of the system and allow for an increase of the time step to typically 2 fs.[50,51] Whether the particle motion represents the physically relevant dynamics of the system depends directly on the quality of the force field.

4.25.3   Statistical Mechanics

According to the ergodic hypothesis[52] one can simulate a single molecule with its surroundings for a period of time and get time-averaged molecular properties that approach the experimentally measurable ensemble averages. This means that from a simulation of a system in time, we can get conformations that correspond to a thermodynamic ensemble or state point. In a simulation that is performed according to eqns [5] and [6] the total energy of the system, H or E, will be conserved as well as the number of particles, N, and the volume, V. This corresponds to the microcanonical (N,V,E) ensemble. Other relevant ensembles are the canonical ensemble (N, V, T), in which the temperature, T, rather than the energy is kept constant and the isothermal–isobaric ensemble (N, p, T), where the pressure p is constant additionally. Simulations corresponding to these ensembles can also be carried out by adding the proper algorithms to the simulation protocol.[7,9,53]

The probability to find a specific configuration of the system defined through the positions and momenta of the (indistinguishable) particles is defined by the phase-space probability $P(\mathbf{r},\mathbf{p})$. In the canonical ensemble one can write the phase-space probability as

$$P_{NVT}(\mathbf{r}, \mathbf{p}) = \frac{e^{-H(\mathbf{r},\mathbf{p})/k_B T}}{\int \int e^{-H(\mathbf{r},\mathbf{p})/k_B T} d\mathbf{p} \ d\mathbf{r}} = \frac{e^{-H(\mathbf{r},\mathbf{p})/k_B T}}{h^{3N} N! Z(N,V,T)} \tag{7}$$

where k_B is the Boltzmann constant, h is Planck's constant, and $Z(N,V,T)$ is the canonical partition function, defined as

$$Z(N,V,T) = \frac{1}{h^{3N} N!} \int \int e^{-H(\mathbf{r},\mathbf{p})/k_B T} d\mathbf{p} \ d\mathbf{r} \tag{8}$$

In the isothermal–isobaric ensemble the partition function is written as

$$Z(N,p,T) = \frac{1}{V h^{3N} N!} \int \int \int e^{-(H(\mathbf{r},\mathbf{p})+pV)/k_B T} d\mathbf{p} \ d\mathbf{r} \ dV \tag{9}$$

and the phase-space probability also depends on the volume of the system,

$$P_{NpT}(\mathbf{r}, \mathbf{p}, V) = \frac{e^{-(H(\mathbf{r},\mathbf{p})+pV)/k_B T}}{V h^{3N} N! Z(N,p,T)} \tag{10}$$

For any experimentally measurable property Q, that may depend on $\mathbf{r}$ and $\mathbf{p}$, the measured value will be equal to the expectation value of the property over the whole phase-space. For the canonical ensemble we write

$$\langle Q \rangle_{NVT} = \int \int Q(\mathbf{r}, \mathbf{p}) \cdot P(\mathbf{r}, \mathbf{p}) d\mathbf{p} \ d\mathbf{r} = \frac{\int \int Q(\mathbf{r}, \mathbf{p}) \cdot e^{-H(\mathbf{r},\mathbf{p})/k_B T} d\mathbf{p} \ d\mathbf{r}}{\int \int e^{-H(\mathbf{r},\mathbf{p})/k_B T} d\mathbf{p} \ d\mathbf{r}} \tag{11}$$

where angular brackets indicate an ensemble average. The ergodic hypothesis can now be described mathematically by stating that such an ensemble average should in the case of sufficient sampling be equal to the long-time average of the time dependent quantity Q.

$$\langle Q \rangle_{NVT} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^{\tau} Q(\mathbf{r}(t), \mathbf{p}(t)) dt \tag{12}$$

A major limitation in molecular dynamics simulations is often the fact that for many biologically interesting properties the timescales reached in computer simulations are still not long enough to obtain convergence to this limit.

4.25.4   Calculation of Free Energy

A key equation from statistical mechanics links the Helmholtz free energy, A, of a system to the canonical partition function,

$$A(N,V,T) = -k_B T \ln Z(N,V,T) \tag{13}$$

Similarly the Gibbs free enthalpy, G, can be calculated from the isothermal–isobaric partition function,

$$G(N,p,T) = A(N,V,T) + pV = -k_B T \ln Z(N,p,T) \tag{14}$$

From these equations we can see that all one needs to do in order to calculate the free energy of the system is to calculate the integral over all phase space in eqns [8] or [9]. Or in terms of a simulation, one needs to sample until the

system has visited all of the conformational or configurational space. If we now recall that $\mathbf{r}$ and $\mathbf{p}$ in these integrals represent $3N$-dimensional vectors, it becomes clear that we can not hope to calculate the absolute free energy of a system containing more than a handful of particles.[54]

4.25.4.1   Overlapping Ensembles

Fortunately, in order to understand the behaviour of a system one does not need to know the absolute free energy, but rather the change in free energy corresponding to some process or state change. For example in the case of solvating a small molecule in water, we are interested in the free energy of the solvated system relative to the molecule in the gas phase. In the case of a ligand binding to a protein we compare the free energy in the bound state to that of the ligand in solution and in the case of DNA base pairing we compare the free energy in the paired state to that of a situation in which both bases are unpaired. For the relative free energy between two states A and B of the system we can write,

$$
\begin{aligned}
\Delta A_{BA} &= A_B - A_A = -k_B T \ln \frac{Z_B(N,V,T)}{Z_A(N,V,T)} \\
&= -k_B T \ln \frac{\int \int e^{-H_B(\mathbf{r},\mathbf{p})/k_B T} d\mathbf{p}\, d\mathbf{r}}{\int \int e^{-H_A(\mathbf{r},\mathbf{p})/k_B T} d\mathbf{p}\, d\mathbf{r}} \\
&= -k_B T \ln \frac{\int \int e^{-(H_B(\mathbf{r},\mathbf{p})-H_A(\mathbf{r},\mathbf{p}))/k_B T} e^{-H_A(\mathbf{r},\mathbf{p})/k_B T} d\mathbf{p}\, d\mathbf{r}}{\int \int e^{-H_A(\mathbf{r},\mathbf{p})/k_B T} d\mathbf{p}\, d\mathbf{r}} \\
&= -k_B T \ln \left\langle e^{-(H_B - H_A)/k_B T} \right\rangle_A
\end{aligned}
$$

[15]

where Z_A and Z_B indicate the canonical partition function of the system while it is in the corresponding state. We have used the short notation H_A for $H_A(\mathbf{r},\mathbf{p})$ and likewise for H_B. It is clear that the free energy of state B relative to that of state A can be calculated from an ensemble average of the Boltzmann factor, $e^{-(H_B-H_A)/k_B T}$ at state A. Equation [15] is a formulation of the well-known perturbation formula, due to Zwanzig.[17] Making use of ergodicity, the ensemble average in this equation can be obtained from the time average of the exponential from a simulation at state A. The sampling problem that was hinted at in the previous section, is reduced for the calculation of a relative free energy. One no longer needs to sample all of conformational space, but only those parts of conformational space that are relevant to and energetically different for both states A and B. If the conformational space relevant to state B is very different to that of state A, it is not likely that a simulation at state A will yield an accurate estimate of the ensemble average. Only if the ensembles of states A and B show considerable overlap, can one hope to reach convergence in the ensemble average within reasonable time.

Graphically, this is depicted in the top panel of **Figure 1**. From a simulation of state A, that samples configurations in the left-hand side of a one-dimensional conformational space, one will not expect to accurately reproduce the ensemble average for the free energy difference with state B, which corresponds mainly to the right hand side of the conformational space. For a state B′ (dashed curve) the ensemble average in eqn [15] is more likely to reach convergence.

## 4.25.4.2   Thermodynamic Cycles

As a first attempt to calculate free energy differences one can try to formulate the problem in terms of free energies between states which are as close as possible in conformational space. This can for instance be done by recognizing the fact that a computer simulation is not limited to physical processes, but can also be used to calculate free energy differences for nonphysical processes. Knowing that the free energy is a state function and does not depend on the pathway connecting the two states of interest, one can try to reformulate the problem using a thermodynamic cycle.[55] Consider for example a case where one is interested in the free energy of binding of a compound to a receptor. For pharmaceutical purposes, one is often mostly interested in the free energy of binding relative to another compound. In **Figure 2** the binding processes for compounds 1 and 2 to a receptor are depicted. The relative free energy of binding can be calculated as

$$
\Delta \Delta A_{bind}(2,1) = \Delta A_{bind}(2) - \Delta A_{bind}(1)
$$

[16]

One can imagine that the ensembles corresponding to the end-states of the binding processes are extremely different since the compound is in the free state physically not even near the protein. Simulations that cover the spontaneous binding of a compound to a protein are still very much beyond currently accessible timescales. From the fact that the free energy is a state function, however, one can also calculate the same relative free energy of binding from the difference between two mutation free energies,

$$
\Delta \Delta A_{bind}(2,1) = \Delta A_{21}(\text{bound}) - \Delta A_{21}(\text{free})
$$

[17]

Conformational space
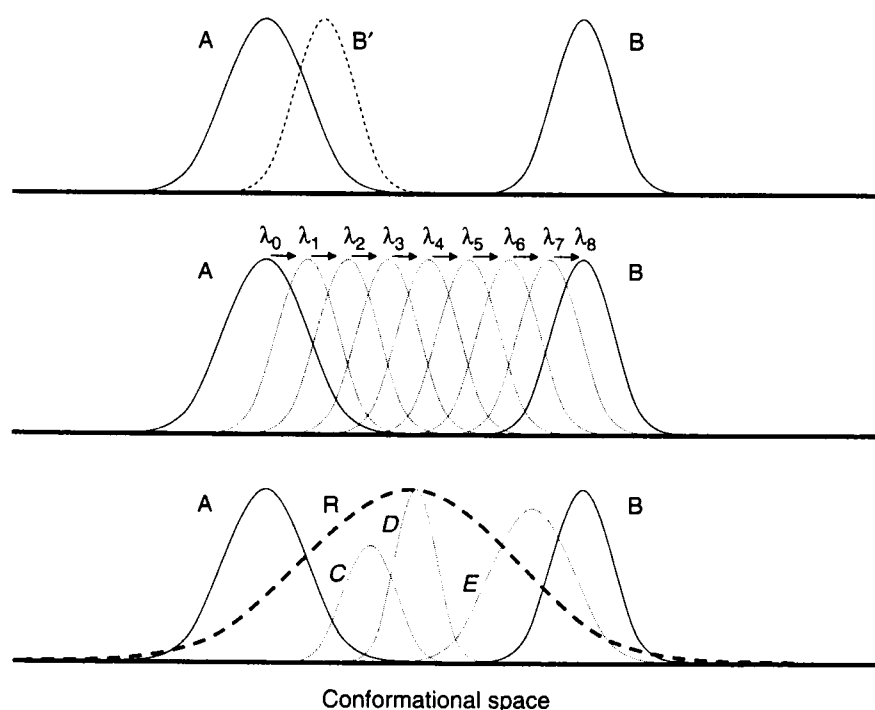
Figure 1 Pictorial representation of the distribution in conformational space for ensembles of different states. The free energy difference between states A and B cannot be computed directly from eqn [15] in case of finite sampling (top panel). Multistep approach (middle panel). One-step from an unphysical reference state (lower panel). See text for details.
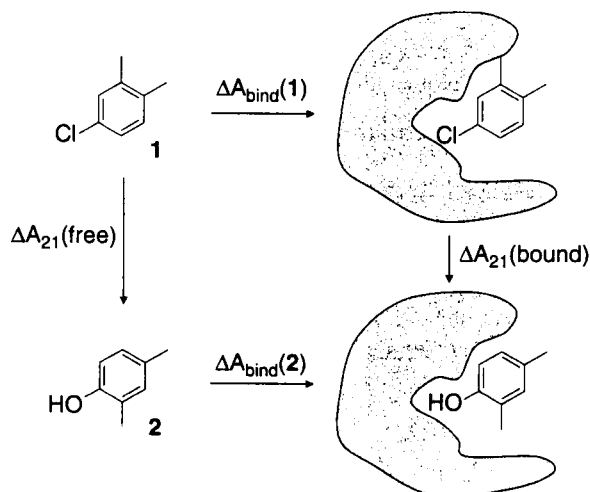
Figure 2 Thermodynamic cycle for the calculation of the relative free energy of binding of compounds **1** and **2** to a common receptor.

Even though the mutation free energies, as indicated by the vertical arrows in **Figure 2** do not correspond to any direct physical process, the end-states of the respective 'processes' can be expected to be much closer in conformational space. By reformulating the problem, the same relative free energy can be calculated by comparing states A and B in the top panel of **Figure 1** that are closer in conformational space.

Unfortunately, a reformulation of the problem does often still not bring the ensembles between which the free energy difference needs to be calculated close enough to expect the ensemble average in eqn [15] to converge within reasonable time. The conformational space relevant to compound 1 may still be very different from the conformational space relevant to compound 2. In the following section different approaches will be discussed to calculate the free energy difference between two states A and B that do not overlap in their conformational spaces.

4.25.4.3   Free Energy Perturbation

The middle panel of **Figure 1** suggests one possibility to obtain the free energy difference between states A and B, commonly known as the free energy perturbation method (FEP).[14,19] The free energy difference that will not converge can be divided into smaller steps for which the free energy change can be calculated individually. This requires the definition of a system that is in an intermediate state between A and B, which can be done using the coupling parameter approach in which the Hamiltonian is written as a function of a coupling parameter, λ, e.g.,

$$H(\mathbf{r}, \mathbf{p}, \lambda) = (1 - \lambda) \cdot H_A(\mathbf{r}, \mathbf{p}) + \lambda \cdot H_B(\mathbf{r}, \mathbf{p}) \qquad [18]$$

Using this definition we ensure that at $\lambda = 0$, $H(\mathbf{r}, \mathbf{p}, 0) = H_A(\mathbf{r}, \mathbf{p})$ and at $\lambda = 1$, $H(\mathbf{r}, \mathbf{p}, 1) = H_B(\mathbf{r}, \mathbf{p})$. At intermediate λ-values the Hamiltonian is a linear combination of the two. A coupling parameter between different states of a system was already used by Kirkwood in 1935[16] and goes back to the work of De Donder 8 years earlier.[56] Different functional forms and pathways interpolating the two end-state Hamiltonians can be used as well.[57–59] As depicted in the middle panel of **Figure 1**, the free energy between states A and B can now be calculated as the sum of $N_\lambda - 1$ free energy differences between adjoining λ-points

$$\Delta A_{BA} = \sum_{i=0}^{N_\lambda-2} A_{\lambda_{i+1}} - A_{\lambda_i} = \sum_{i=0}^{N_\lambda-2} -k_B T \ln \left\langle e^{-(H(\mathbf{r},\mathbf{p},\lambda_{i+1}) - H(\mathbf{r},\mathbf{p},\lambda_i))/k_B T} \right\rangle_{\lambda_i} \qquad [19]$$

This requires a simulation to be carried out at every (except for one) of the N_λ λ-values, in order to calculate the ensemble average at that λ-point.

4.25.4.4   Thermodynamic Integration

A slightly different approach is taken in the thermodynamic integration method (TI).[16] It also makes use of the coupling parameter approach, as in eqn [18]. The free energy difference can be calculated as the integral from $\lambda = 0$ (state A) to $\lambda = 1$ (state B) of the derivative of the free energy A with respect to λ,

$$\Delta A_{BA} = \int_0^1 \frac{dA(\lambda)}{d\lambda} d\lambda \qquad [20]$$

Direct application of eqn [13] yields for the derivative

$$
\begin{aligned}
\frac{dA(\lambda)}{d\lambda} &= \frac{d}{d\lambda}\left(-k_B T \ln \frac{1}{h^{3N} N!} \int \int e^{-H(\mathbf{r},\mathbf{p},\lambda)/k_B T} d\mathbf{p}\ d\mathbf{r}\right) \\
&= -k_B T \frac{\frac{d}{d\lambda}\int \int e^{-H(\mathbf{r},\mathbf{p},\lambda)/k_B T} d\mathbf{p}\ d\mathbf{r}}{\int \int e^{-H(\mathbf{r},\mathbf{p},\lambda)/k_B T} d\mathbf{p}\ d\mathbf{r}} \\
&= \frac{\int \int \frac{\partial H(\mathbf{r},\mathbf{p},\lambda)}{\partial \lambda} e^{-H(\mathbf{r},\mathbf{p},\lambda)/k_B T} d\mathbf{p}\ d\mathbf{r}}{\int \int e^{-H(\mathbf{r},\mathbf{p},\lambda)/k_B T} d\mathbf{p}\ d\mathbf{r}} \\
&= \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda \qquad [21]
\end{aligned}
$$

being exactly the ensemble average of the derivative of the Hamiltonian with respect to λ from an ensemble generated at that λ-value. The free energy difference can be calculated from a numerical integration of this ensemble average obtained from N_λ simulations at different λ-values. Like for the free energy perturbation approach, several simulations at intermediate values are required. The approach of sampling different areas of conformational space at different λ-values, is depicted in **Figure 3** in the upper left panel. In this plot, state A in **Figure 1** corresponds to the lower part of the one-dimensional conformational space, while state B corresponds to the upper part of the vertical axis.

4.25.4.5   Slow Growth

The slow growth approach can be seen as a modification of the FEP or TI approaches. Historically, TI calculations were originally carried out in a slow growth manner. Instead of simulating at discrete λ-points, the value of λ is increased at every time step such that $\lambda = 0$ (state A) at the start of the simulation and $\lambda = 1$ (state B) at the end of the simulation. This is depicted by the bold line in the upper right-hand corner of **Figure 3**. It can easily be shown that if eqn [18] is
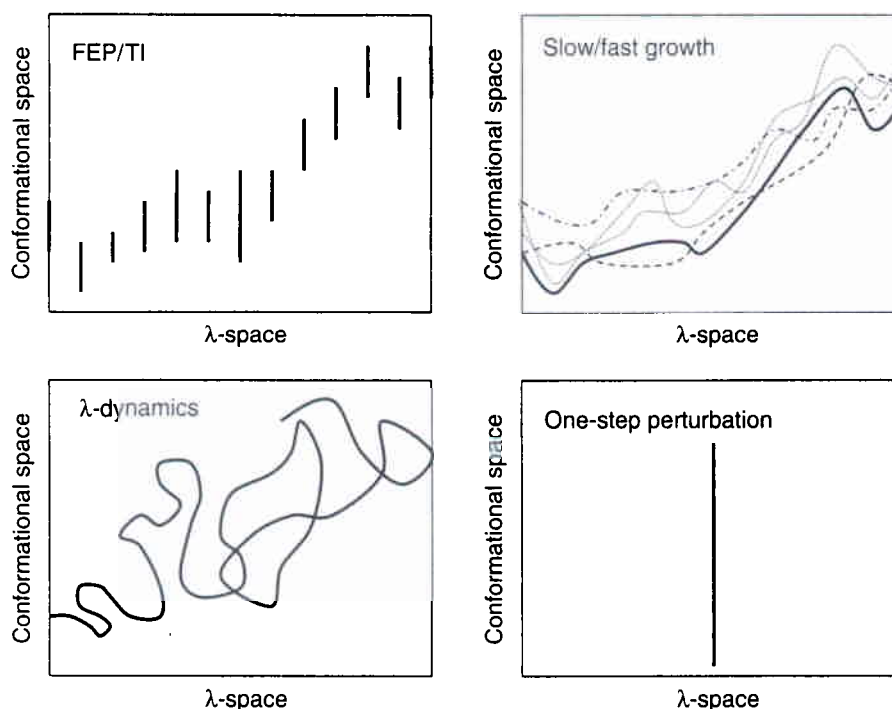
Figure 3   Pictorial representation of motion in conformational and λ-space for different approaches to free energy calculations. See text for details.

used to interpolate the Hamiltonian between states A and B, applying the perturbation formula [19] or the thermodynamic integration formula [20] will yield exactly the same estimate of the free energy difference. For different λ-dependencies of the Hamiltonian, the different approaches might result in different numerical results.

The major drawback of the slow growth approach is that the ensemble averages at every λ-value in eqns [19] or [21] are estimated from a single configuration. Obviously, this will be a very poor estimate if one considers the ergodic long time limit of the average in eqn [12]. In other words, because the λ-value is continuously changing, the system is never allowed to reach equilibrium. The motion of the particles will never be able to adapt to the new, slightly changed Hamiltonian, because the latter is constantly changing. Only if the transition from state A to state B is performed very slowly, using very long simulations, will the resulting free energy estimate converge toward the actual value, even though the conceptual error of approximating an ensemble average by a single value will remain even under such conditions. A slow growth estimate of the free energy difference is fundamentally a nonequilibrium estimate.[30] For this reason, the pure slow growth approach is no longer applied on a large scale. However, it has found a new application in the so-called fast growth approach, which will be discussed in Section 4.25.4.7.

4.25.4.6   λ-Dynamics

Reminiscent of a slow growth approach is a method that is depicted in the lower left-hand panel of **Figure 3**. In the λ-dynamics approach,[60] the value of λ is also continuously changing, but is treated as another degree of freedom in an extended Hamiltonian. The system is free to dynamically optimize the value of λ during the course of a simulation, possibly overcoming barriers that would occur in a slow growth simulation. The free energy is obtained by calculating the potential of mean force as a function of λ. To make sure that the system gradually moves from λ = 0 to λ = 1, and that all values of λ are sampled sufficiently, λ is often restrained to stepwise increasing values, through an umbrella sampling technique.[61] As in the FEP and TI approaches, several simulations at different λ-values will thus be required to obtain the free energy difference. The λ-dynamics method is particularly used to have one or several coupling parameters scale between different ligands bound to a common receptor. From the average value of the λ-values, the preference of the system for one of the ligands can be determined. This raises the question of what is the physical meaning of a system that prefers a ligand that consists for 80% of compound 1 and 13% of compound 2 and 7% of compound 3. A similar approach is the chemical Monte Carlo approach, where a single ligand is bound to the protein at all times, but the ligand changes character based on a random, Monte Carlo method.[62–64]

4.25.4.7   Fast Growth

Recently, the slow growth method to obtain free energy differences has received renewed attention thanks to the work of Jarzynski.[65,66] It was shown that from a canonical collection of N_A nonequilibrium free energy estimates $\Delta A'$, such as those obtained from slow growth simulations, one can obtain an equilibrium estimate for the free energy from

$$\Delta A_{BA} = -k_B T \ln\frac{1}{N_A} \sum_{i=1}^{N_A} e^{-\Delta A'_{BA}/k_B T} \qquad [22]$$

The upper right-hand corner of **Figure 3** demonstrates this approach based on slow growth simulations. Many different slow growth simulations from a canonical ensemble of starting configurations lead to many different trajectories. The required sampling of conformational space, represented by the vertical lines in the upper left-hand corner of **Figure 3** is not obtained sequentially at discrete λ-points, but is acquired as more slow growth trajectories are sampled. This procedure has become known as the fast growth method.[67,68] Even though the Jarzynski equation is often applied to slow growth estimates to obtain the true free energy difference, it is not restricted to it. In principle it can be applied to any canonical ensemble of (nonequilibrium) free energy estimates.

4.25.4.8   One-Step Perturbation

The methods to calculate the free energy difference between states A and B in **Figure 1** that have been discussed so far were all aimed at gradually changing the system A into system B, be it in a limited number of discrete steps, or in a continuous, but nonequilibrium manner. The one-step perturbation method takes a different approach. It is based on the free energy perturbation formula [15], but tackles the problem of nonoverlapping ensembles by generating a single, broad ensemble, as displayed in the lower panel of **Figure 1**. The underlying idea is that we use a single, not necessarily physical, reference state R of which the ensemble shows overlap with the ensembles of both states A and B. For each of these states the free energy difference from the reference state can be calculated directly from a single simulation of the reference state using eqn [15], or

$$\Delta A_{AR} = A_A - A_R = -k_B T \ln\left\langle e^{-(H_A - H_R)/k_B T} \right\rangle_R \qquad [23]$$

giving rise to the name one-step perturbation.[69] The free energy difference between states A and B is then simply

$$\Delta A_{BA} = \Delta A_{BR} - \Delta A_{AR} \qquad [24]$$

One might argue that we have now only shifted the problem from obtaining overlap between states A and B to generating a reference state that covers both of these states simultaneously. A solution to this problem can be found by recognizing that the formulae leading to ΔA_{BA} in eqn [24] do not require the reference state R to actually correspond to a physically meaningful state. In the example of calculating the relative free energies of two species, it is not required that the reference state corresponds to a real molecule itself. In fact, a considerable broadening of the ensemble can be obtained by making some of the atoms in the reference state 'soft.'[70,71] That is, let the interaction of these atoms level off to a finite value at the origin, rather than to an infinitely high repulsion. See for instance **Figure 3** in Oostenbrink and van Gunsteren[72] in which it is shown that such a reference state samples both configurations in which an atom is present as well as configurations that closely resemble situations where the atom is not there at all. The ensemble that is obtained by simulating an nonphysical reference state shows overlap with both the ensemble that corresponds to a simulation of the real state A and the ensemble of the real state B. In fact, the method is not restrictive to states A and B. As indicated in the lower panel of **Figure 1**, the reference state can also show overlap between additional states of interest, C, D, and E. From the single simulation of the reference state, free energies between all these states can be obtained. In addition, structural information about the real states can be obtained by a close inspection of those configurations from the simulation of R that contribute most to the ensemble average in eqn [23]. Obviously these configurations will belong to the most relevant configurations for that particular real state.

Equation [23] can be seen as an application of the Jarzynski eqn [22] with an infinitely fast sampling of the individual free energy estimates, $\Delta A' = H_A - H_R$, taken at every configuration of the simulation of state R.

4.25.4.9   Other Methods

In this section several methods to obtain free energies will be discussed which have a more empirical background. These commonly calculate the free energy difference between states A and B from simulations at these states, either by estimating the absolute free energies for both end-states or by computing specific energetic differences between

the states.[73,74] The advantage of such approaches is that one only needs to simulate systems with physical relevance, as opposed to the previously discussed methods that all involve simulations of nonphysical states of the system. However, compared to the one-step approach in which the number (few) of simulations is independent of the number of real states, here this number of simulations is at least equal to the number of real states.

The first method that will be discussed is the linear interaction method (LIE) due to Åqvist et al.[75,76] It has found widespread application to calculations of ligand–protein complexation free energies[77] and of free energies of solvation for small compounds.[78] The free energy difference between the states A and B is written as

$$\Delta A_{BA} = \alpha \, \Delta \langle V^{vdw} \rangle_{BA} + \beta \, \Delta \langle V^{el} \rangle_{BA} + \gamma \qquad [25]$$

where $\Delta \langle V^{vdw} \rangle_{BA}$ is the difference of the ensemble average of the van der Waals energy in states B and A, and $\Delta \langle V^{el} \rangle_{BA}$ is the difference of the ensemble average of the electrostatic energy in states B and A. Often only the contribution to these energies that are expected to change most are taken into account. In the case of a ligand binding to a protein, where state A represents the ligand in water and state B the ligand bound to the protein, this would typically include only the interaction energy of the ligand with its surroundings. α, β, and γ are empirical parameters that are obtained by fitting calculated free energy differences to experimental data for a set of compounds. In some cases γ depends on the solvent accessible surface area (SASA) of the molecule of interest. This means that a learning set of experimental data to obtain these empirical parameters is required to calculate free energies for similar systems. From linear response theory, a theoretical value for β of 1/2 can be derived, which seems to hold well for ionic compounds, but is less accurate for dipolar and potential hydrogen bonding compounds.[79]

Another popular method is aimed at estimating the absolute free energy for the end-states of a simulation (excluding the kinetic energy). In the MM-PBSA method, the free energy is estimated from four contributions, using[80]

$$\langle A \rangle = \langle V^{MM} \rangle + \langle V^{PB} \rangle + \langle V^{SA} \rangle - TS^{MM} \qquad [26]$$

V^{MM} is the molecular mechanics potential energy as in eqn [1], V^{PB} is the electrostatic solvation (free) energy of the solute, as calculated from a numerical solution of the Poisson–Boltzmann equation, V^{SA} is a nonpolar contribution to the solvation (free) energy, estimated from the SASA of the solute, T is the temperature, and S^{MM} is the configurational solute entropy, which can be estimated from a quasiharmonic analysis or a normal-mode analysis of the molecular dynamics simulation. The accuracy of the MM-PBSA method is generally lower than conventional free energy perturbation calculations due to the numerous approximations and assumptions that go into the calculation of especially the last three terms. However, the computational gain is significant and impressive correlations with experimental free energies are often reported.

Empirical free energy scoring functions are popular for computer-aided molecular design purposes.[81] In general, the free energy of ligand binding is estimated from an equation involving several terms and empirical parameters. For example, such terms can depend on the presence of hydrogen bonds or hydrophobic contacts between the protein and the ligand, or can be related to the overall partition coefficient ($\log P$) of the ligand, the solvent-accessible surface of the ligand or size of the protein–ligand interface. The parameters are usually obtained from a regression analysis on a number of protein–ligand complexes with known binding affinity. Scoring functions are frequently used in automated ligand design programs,[82,83] or by programs that dock a given ligand into the binding site of a protein.[84–86] Typically, only a single or a few rigid conformations of the protein (or even the ligand) are considered, which severely limits the reliability and the theoretical support of these approaches. On the other hand, examples are known where the conformational changes from molecular dynamics simulations seem to have a limited influence on the estimated free energy values.[87]

4.25.5   Example: The Estrogen Receptor

This section describes examples of usage of the methods that were introduced above. Molecular dynamics simulations have proven useful for drug design purposes on many different targets.[88–103] Here, we will focus on the ligand-binding domain of the estrogen receptor. The estrogen receptor is a member of the nuclear hormone receptors and plays a key role in the growth, development, and maintenance of a diverse range of tissues. It consists of an N-terminal DNA-binding domain, a ligand-binding domain, and a C-terminal activation domain. Upon binding of an agonist to the ligand-binding domain a conformational change in the estrogen receptor takes place, allowing it to homodimerize and subsequently to translocate to the nucleus. Here, the DNA-binding domain directly interacts with response elements on the DNA, thereby activating or repressing transcription.[104] Apart from the physiological effects, the estrogen receptor is involved in a range of diseases such as breast cancer, osteoporosis, endometrial cancer, and prostate hypertrophy.[105]

Naturally occurring estrogens are steroid hormones, such as the endogenous ligand 17β-estradiol. Apart from this endogenous ligand, the estrogen receptor is known to show affinity for a wide range of structurally diverse compounds, including synthetic estrogens, phytoestrogens, pesticides, polychlorinated biphenyls (PCBs), and polycyclic aromatic hydrocarbons (PAHs).[106-110] PAHs are products of incomplete combustion of fossil fuels, wood, and other organic matter and as such are ubiquitous in the environment. Oxidative biotransformation by cytochrome P450 introduces hydroxyl groups, which results in structures that can mimic estradiol.[111-113] Everyday foodstuffs are another source of estrogenlike compounds. Some foods are known to contain a wide range of phytoestrogens, such as the isoflavonoids genistein and daidzain that show estrogenic activity in vitro and in vivo. The most important human sources of these isoflavonoids are soybeans and soybean products.[114] Although the production and use of polychlorinated biphenyls was banned in the late 1970s, they have still been observed in all kinds of tissues and species.[115-117] Again, one of the metabolic pathways of the PCBs involves the hydroxylation of (one of the) aromatic rings,[118] after which the affinity for the estrogen receptor is altered.[119]

Several crystal structures of the estrogen receptor ligand-binding domain in complex with various agonists and antagonists are available from which it becomes clear that the ligand occupies a deeply buried, largely hydrophobic cavity within the ligand-binding domain. In most structures there is no clear path or channel connecting this cavity with the exterior of the protein, in contrast to typical protein–ligand complexes. In addition, the structure of an antagonist bound to the estrogen receptor ligand-binding domain shows that a bulky side chain moiety of the antagonist serves to prevent the formation of the active tertiary structure of helices 11 and 12, known to be crucial for activity.[120,121]

The overall topology of the estrogen receptor ligand-binding domain, the wide structural diversity of known agonists, its physiological relevance, and the availability of crystal structures of the target make the estrogen receptor an ideal candidate to apply and develop computational methods for the prediction of binding affinities, to be used for drug design purposes. The following sections will describe some of our experiences of molecular dynamics simulations and free energy calculations of the estrogen receptor ligand-binding domain, the prediction and structural interpretation of binding affinities, and the design of a new ligand binding to the estrogen receptor.[72,101,122,123]

4.25.5.1   Simulation of Known Ligands

A detailed understanding of the structure and dynamics of protein–ligand complexes and the interactions between proteins and ligands has proven to be very helpful in designing new potent drugs. Producing crystallographic structures of many different ligands is still very time-consuming and only provides structural information about the complex. In order to capture and understand the flexibility of protein and ligand, molecular dynamics simulations are especially well suited. Such simulations have been performed on the estrogen receptor ligand-binding domain, complexed with the synthetic estrogen diethylstilbestrol (DES) using the crystal structure 3ERD[124] from the RCSB Protein Data Bank.[125] This crystal structure contains residues Ser305 to Leu549, while the side chains of 18 residues were not resolved and modeled according to standard side chain configurations from the GROMOS96 simulation package.[39,126] The estrogen receptor ligand-binding domain–DES complex was solvated in a periodic rectangular box of $520 \, nm^3$, filled with 15824 explicit SPC water molecules.[127] After proper equilibration the system was simulated for 1 ns, at a constant temperature of 298 K and a constant pressure of 1 atm.

Before exploring the protein–ligand dynamics and the interactions between them, the simulation needs to be validated in terms of energetic and structural stability. The first is done by ensuring that the potential and kinetic energy have converged to reasonably constant values, while the latter is often expressed in terms of the atom-positional root-mean-square deviation from the initial (crystal) structure.[122] The presence and stability of secondary structure elements over the course of the simulation can be calculated as well, where it should be noted that these are not primary experimental data. The crystal structure from which the secondary structure elements are inferred is merely a model that represents the primary experimental data (electron density or rather structure factor amplitudes) best. One could also compare to primary structural data, such as NMR derived upper bounds to proton–proton distances (or rather NOE intensities).

A well-established estrogenic pharmacophore model consists of two hydroxyl groups, at a distance of 1.0 to 1.2 nm. At least one of the hydroxyl groups is bound to an aromatic ring and in between the two hydroxyl moieties, the estrogen is mainly hydrophobic.[111] From the crystal structures these pharmacophoric features could be explained by a hydrogen bonding network consisting of Glu353, Arg394, a water molecule, and the aromatic hydroxyl group of the ligand on the one (proximal) side and a hydrogen bond between His524 and the other hydroxyl group on the other (distal) side (**Figure 4**). From the simulations, the occurrence of protein–ligand hydrogen bonds was monitored. This analysis revealed that the hydrogen bonding interactions are highly dynamic; hydrogen bonds are formed and broken
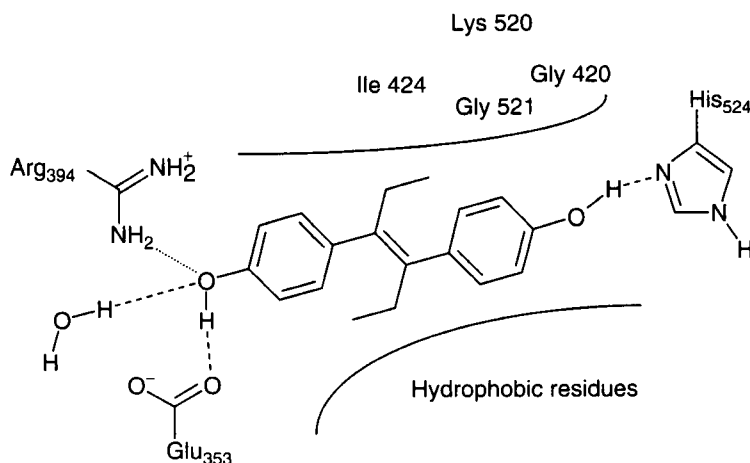
Figure 4 Schematic representation of ligand diethylstilbestrol (DES) bound to the estrogen receptor active site.
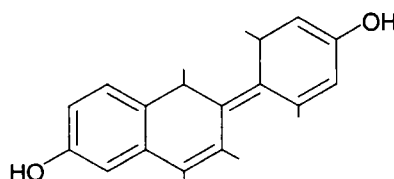
Figure 5 Reference compound (REF) used in the free energy calculations. Soft atoms are represented as gray circles.

continuously over time. While none of the hydrogen bonds at the proximal side of the molecule was present for more than 60% of the time, this hydroxyl group is always involved in at least one hydrogen bond, and forms on average 1.5 hydrogen bonds. The distal hydroxyl group forms a hydrogen bond to His524 for 97% of the simulation time.

During protein solvation in the computational box, two noncrystallographic water molecules were added near one of the ethyl groups of the DES in a hydrophobic cavity. These waters leave the binding site past His524 and toward the charged moiety of Lys520. No other water molecules enter the binding site, and the crystallographic water molecules found in the hydrogen bonded network remain trapped next to the ligand over the entire simulation. Once the added water molecules have left the binding site (after 300 ps), the ethyl group of DES shows much more flexibility in the cavity as was observed from the fluctuations of the dihedral angle. This is relevant information that can be used in the design of alternative ligands.

4.25.5.2   Calculation of Binding Free Energies

Thermodynamic integration was used to calculate the free energy of ligand binding between DES, estradiol, and genistein. These compounds are quite different to one another and relative binding affinities were obtained that agree well with experimental data. However, for drug design purposes TI is often too inefficient to screen multiple compounds on their binding affinity. In order to estimate the potential toxic effects of modified hydroxylated polyaromatic hydrocarbons through the estrogen receptor, the one-step perturbation method was applied. As a reference state, the compound (indicated with REF) in **Figure 5** was used, in which the indicated atoms were described by the so-called soft-core potential.[70] For these atoms, the repulsive singularity at the origin was absent, allowing surrounding atoms to show considerable overlap with these atoms from time to time. In this way, the broad ensemble in the lower panel of **Figure 1** was obtained, that overlaps with the ensembles that would be generated for the compounds listed in **Figure 6**.

By applying eqn [23] over a 1-ns trajectory of this reference state in the unbound state (solvated in water) and when bound to the protein (solvated in water), the relative free energies of binding that are indicated in **Figure 6** were obtained. From a comparison of computed free energies of binding relative to the reference state to the experimental values, the free energy of binding for the (unphysical) reference state can be estimated as about $-57\,kJ\,mol^{-1}$. It can be seen from **Figure 7** that the calculated relative free energies of binding compare reasonably well to the
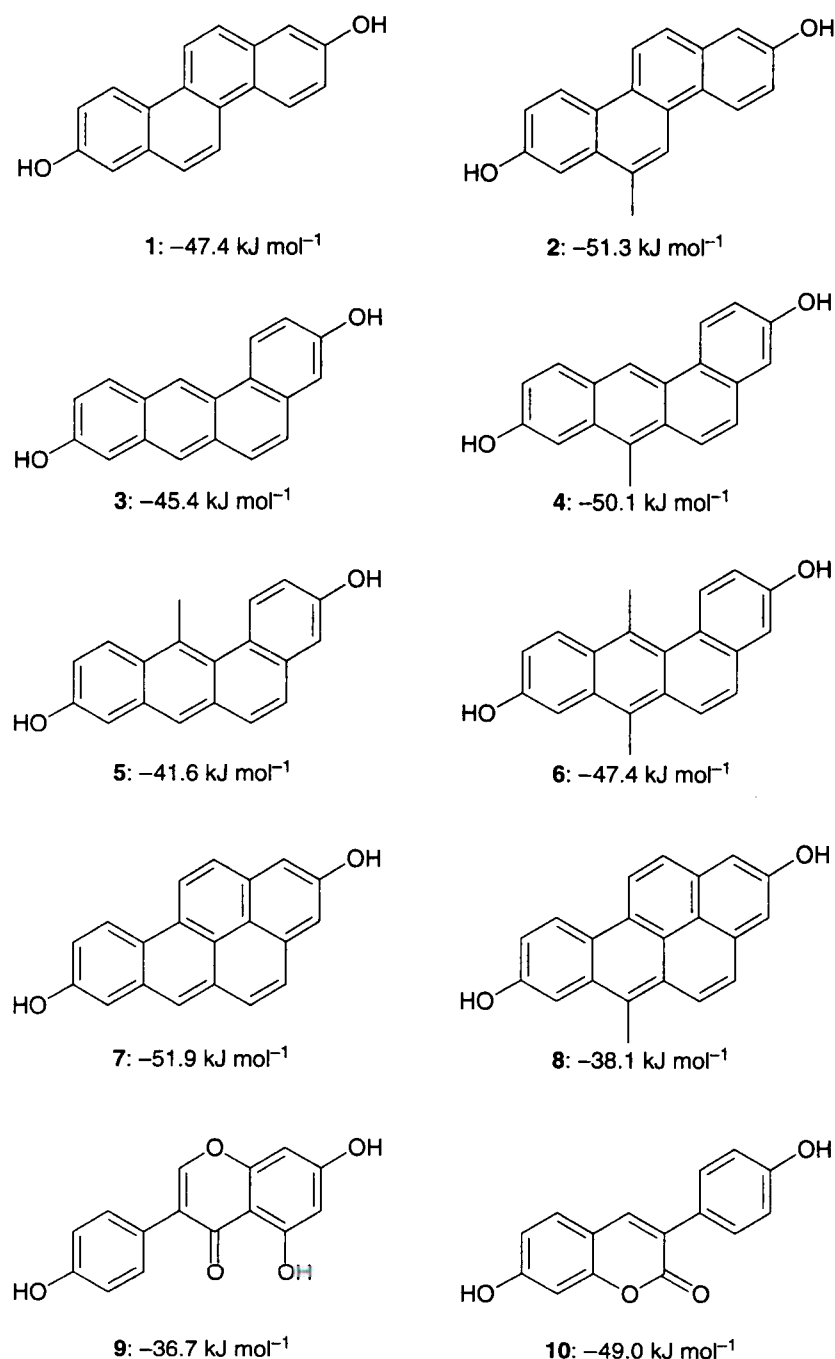
Figure 6 Compounds for which the relative free energy of binding was calculated. Indicated are free energies of binding using an estimated binding affinity of the reference compound of $-57 \, \text{kJ mol}^{-1}$.

experimental values, with an absolute mean error of $1.7 \, \text{kJ mol}^{-1}$. The maximum deviation occurs for compound **4** and is still only $3.1 \, \text{kJ mol}^{-1}$.

It is important to note the efficiency gain of the one-step perturbation method as compared to the other methods described in Section 4.25.4. Here, nine relative free energies of binding were calculated from two 1 ns simulations. Alternative methods, such as TI or FEP, would require at least eight sets of simulations, each consisting of typically 10–20 simulations at different intermediate λ-values of several hundreds of picoseconds.

4.25.5.3   Structural Interpretation

Apart from the efficiency of the free energy calculation obtained at the expense of a somewhat reduced accuracy, an important asset of the one-step perturbation method is that it allows for an interpretation of the results at a molecular
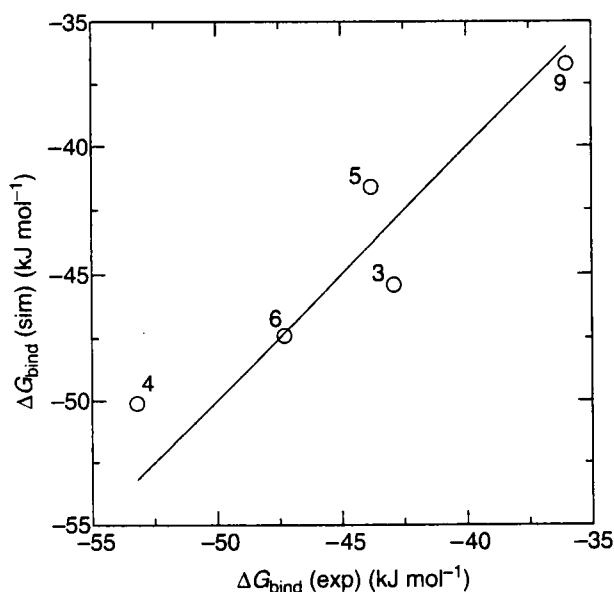
Figure 7 Graphical comparison of experimental and calculated results for the binding free energies of the indicated ligands.

basis. From the trajectory of the reference state, one can extract those configurations of the protein–ligand complex that contribute most to the ensemble average in eqn [23]. The configurations that show the largest Boltzmann probability can then be compared for different real ligands and a structural explanation for different free energy values can be obtained.

Consider for example the polyaromatic hydrocarbons and calculated free energies of binding in **Figure 6**. It is interesting to note that the addition of a methyl to structures **1, 3,** and **5** to obtain structures **2, 4,** and **6** is calculated to improve the binding free energy by about 4–6 kJ mol^{-1}, while introducing a methyl group in the same location in structure **7**, yielding **8**, reduces the affinity by 14 kJ mol^{-1}. It can be expected that three-dimensional quantitative structure–activity relationship (3D-QSAR) methods that do not take the protein structure explicitly into account will have a hard time predicting such nonadditive behaviour. By overlaying the protein–ligand conformation that contributes most to the free energy of binding for compounds **4** and **8** (**Figure 8**), a structural explanation can be given. In compound **8** (in green), all soft atoms in the reference compound are real atoms, forcing an unfavorable interaction of the methyl group with a cysteine (in orange) and leading to an increased binding free energy. For compound **4** (in red), however, two soft atoms in the reference state are not there in the real compound. In this case, the most favorable conformation is one in which the soft atoms show considerable overlap with an Arg side chain (in blue), allowing the methyl group to shift away from the cysteine and fit snugly into a niche in the binding site, reducing the binding free energy as compared to compound **3**.

For rational drug design purposes it is not only important to predict accurately relative binding affinities, but also to understand why one compound interacts more favorably with a receptor or solvent than another compound. In the example described here, we have come to a detailed understanding of the subtle differences between different compounds, and of the details that could be used in the development of novel ligands interacting with the receptor.

4.25.5.4   Drug Design

So far, the example that we have described has been mainly involved with the interpretation of the structure and dynamics of ligand–protein complexes and the reproduction of experimental binding free energies. The model of the reference compound in **Figure 5** has also been used to rationally design a new compound binding to the ligand-binding domain of the estrogen receptor. Experimentally, accurate screening of estrogenic activity in, e.g., environmental samples can be performed using the phytoestrogen coumestrol.[128] This fluorescent compound has the remarkable property that its fluorescence is increased when bound to the estrogen receptor, allowing for a competitive screening of other compounds driving coumestrol away from the binding site, which can be observed directly as a reduction of the fluorescence. For practical purposes, there was a demand for additional fluorescent compounds with a relatively high affinity to the estrogen receptor, but preferably fluorescing at slightly different wavelengths.
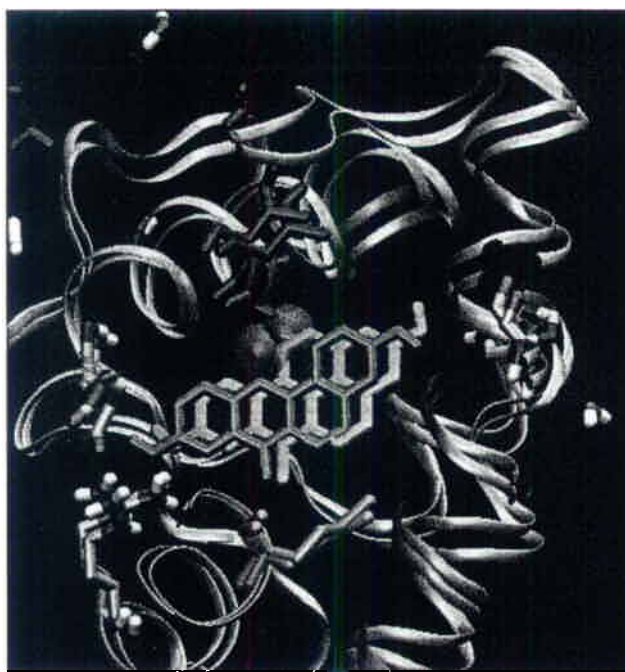
Figure 8 The most contributing configuration of the reference compound for the case when the real compound is compound **4** (in red) and for the case when the real compound is compound **8** (in green) are drawn. In dark blue Leu347 and Thr348 are indicated that show overlap with the soft atoms in the red structure (transparent spheres), allowing it to shift its methyl group away from Met389 (in orange). Protein backbone in cyan. Interacting residues Glu353, Arg394, and His524 in stick representation (compare **Figure 4**).

Based on the fact that coumarines are known to be highly fluorescent and using the knowledge that was obtained in the examples described above, compound 10 in Figure 6 was proposed. Using the estimated free energy of binding of the reference compound of $-57 \, kJ \, mol^{-1}$, the binding free energy for compound 10 was predicted to be $-49.0 \, kJ \, mol^{-1}$. The compound was then synthesized[129] and subsequently tested to yield an experimental binding affinity of $-44 \, kJ \, mol^{-1}$, only $5 \, kJ \, mol^{-1}$ from the predicted value. Subsequent experimental analysis revealed that the fluorescence of this compound is also increased upon binding to the protein.

Even though the compound that was designed in this example was not designed as a potential drug, but rather for analytical purposes, it is clear that accurate free energies of binding can be predicted from molecular dynamics simulations. It is also interesting to note that the predicting strength of the method described here does not depend on training a model to known data. Even without the estimated free energy of binding of the reference compound, compound 10 would have been predicted to bind comparably to compound 6. Moreover, because the protein environment is taken into account explicitly, the method can be used to accurately predict binding affinities of compounds with a more hydrophobic character (compounds 1–8) as well as for compounds that have a more hydrophilic scaffold (compounds 9, 10).

Overall, these examples show that the one-step perturbation method is an accurate and efficient means to predict the free energies of binding for a series of compounds. The protein environment is taken into account explicitly allowing for a detailed structural explanation of calculated differences between compounds. Other than the parameters that go into a (relatively standard) force field, no empirical parameters are required and a statistically sound prediction of the binding affinity is obtained.

4.25.6   Conclusions

This chapter has briefly described methods that can be used to perform molecular dynamics simulations of biomolecular systems. For drug design purposes, this will mainly be receptor–ligand complexes. In addition, we have described methods to calculate the free energies of binding on a sound statistical mechanical basis. From the examples in Section 4.25.5, it should have become clear how these methods can be applied in rational drug design.

First, simulations of protein–ligand complexes give insight in the structure and dynamics of the ligand when bound to the protein. A detailed analysis of the protein–ligand interactions reveals the relative importance of different

interaction sites and allows us to predict the required functional groups in the ligand. From the dynamics of the complex and the average shape of the binding site, possible sites of modification to known ligands can be suggested.

Second, we have shown that molecular dynamics simulations can be used to calculate (relative) binding affinities of a series of compounds with reasonable accuracy. The methods described in Sections 4.25.4.3–4.25.4.8 have in common that no experimental data are required to predict the binding affinities, other than the parameters from a common biomolecular force field. With an average absolute error smaller than k_BT ($\sim 2.5\,\text{kJ mol}^{-1}$) the example that was described shows that such a force field can describe the protein–ligand interactions quite accurately.

Third, the big advantage of molecular dynamics simulation and in particular the one-step perturbation approach is that an explanation of the structural basis of the calculated free energies can be obtained at an atomic level. Structural details can be obtained that explain why one compound binds more tightly to the protein than another.

Fourth, the one-step perturbation methodology allows the calculation of tens of thousands[130] to hundreds of millions[131] of relative free energies for not too different ligands from only a handful of simulations.

Finally, we have shown how the knowledge of the protein and ligand structure and dynamics, an understanding of the structural reasons behind ligand binding, and the machinery to calculate binding free energies can be used to rationally propose novel compounds that will bind to the same target and to predict the binding affinity of such compounds.

With current availability of computational power, molecular dynamics simulations of protein–ligand complexes have become feasible. The methods described offer insight in structure and dynamics at a resolution often not accessible experimentally and the advantages of taking the protein environment of ligands into account will be evident from the results presented in this chapter.

References

1. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
2. Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1957**, *27*, 1208–1209.
3. Rahman, A. *Phys. Rev. A* **1964**, *136*, 405–411.
4. Rahman, A.; Stillinger, F. H. *J. Chem. Phys.* **1971**, *55*, 3336–3359.
5. McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585–590.
6. Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, UK, 1987.
7. van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992–1023.
8. Frenkel, D.; Smit, B. *Understanding Molecular Simulation*; Academic Press: San Diego, CA, 2002.
9. Tuckerman, M. E.; Martyna, G. J. *J. Phys. Chem. B* **2000**, *104*, 159–178.
10. Hansson, T.; Oostenbrink, C.; van Gunsteren, W. F. *Curr. Opin. Struct. Biol.* **2002**, *12*, 190–196.
11. Norberg, J.; Nilsson, L. *Q. Rev. Biophys.* **2003**, *36*, 257–306.
12. Sarkisov, G. N.; Dashevsky, V. G.; Malenkov, G. G. *Mol. Phys.* **1974**, *27*, 1249–1269.
13. Mruzik, M. R.; Abraham, F. F.; Schreiber, D. E.; Pound, G. M. *J. Chem. Phys.* **1976**, *64*, 481–491.
14. Mezei, M.; Swaminathan, S.; Beveridge, D. L. *J. Am. Chem. Soc.* **1978**, *100*, 3255–3256.
15. Postma, J. P. M.; Berendsen, H. J. C.; Haak, J. R. *Faraday Symp. Chem. Soc.* **1982**, *17*, 55–67.
16. Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300–313.
17. Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
18. Mezei, M.; Beveridge, D. L. *Ann. NY Acad. Sci.* **1986**, *482*, 1–23.
19. Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431–492.
20. van Gunsteren, W. F. Methods for Calculation of Free Energies and Binding Constants: Successes and Problems. In *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*; van Gunsteren, W. F., Weiner, P. K., Eds.; Escom Science Publishers: Leiden, the Netherlands, 1989, pp 27–59.
21. Straatsma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407–435.
22. Kollman, P. *Chem. Rev.* **1993**, *93*, 2395–2417.
23. van Gunsteren, W. F.; Beutler, T. C.; Fraternali, F.; King, P. M.; Mark, A. E.; Smith, P. E. Free Energy via Molecular Simulation: A Primer. In *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*; van Gunsteren, W. F., Weiner, P. K., Eds.; Escom Science Publishers: Leiden, the Netherlands, 1993, pp 315–348.
24. Lamb, M.; Jorgensen, W. L. *Curr. Opin. Chem. Biol.* **1997**, *1*, 449–457.
25. Mark, A. E. Free Energy Perturbation (FEP). In *Encyclopedia of Computational Chemistry*; John Wiley: New York, 1998, pp 1070–1083.
26. van Gunsteren, W. F.; Daura, X.; Mark, A. E. *Helv. Chim. Acta* **2002**, *85*, 3113–3129.
27. Warshel, A. *Acc. Chem. Res.* **2002**, *35*, 385–395.
28. Brandsdal, B. O.; Österberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Åqvist, J. *Adv. Protein. Chem.* **2003**, *66*, 123–158.
29. Weiner, P. K.; Kollman, P. A. *J. Comput. Chem.* **1981**, *2*, 287–303.
30. Pearlman, D. A.; Kollman, P. A. *J. Chem. Phys.* **1989**, *91*, 7831–7839.
31. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Mertz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
32. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
33. MacKerell, A. D., Jr.; Wiórkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946–11975.
34. MacKerell, A. D., Jr.; Bashford, D.; Bellot, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
35. Momany, F. A.; Rone, R. *J. Comput. Chem.* **1992**, *13*, 888–900.
36. Nemethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 6472–6484.

37. Levitt, M. *J. Mol. Biol.* 1983, *168*, 595–620.
38. Levitt, M.; Hirshberg, M.; Sharon, R.; Daggett, V. *Comput. Phys. Commun.* 1995, *91*, 215–231.
39. van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Vdf Hochschulverlag AG an der ETH Zürich: Zürich, Switzerland, 1996.
40. Schuler, L. D.; Daura, X.; van Gunsteren, W. F. *J. Comput. Chem.* 2001, *22*, 1205–1218.
41. Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* 2004, *25*, 1656–1676.
42. Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* 1988, *110*, 1657–1666.
43. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* 1996, *118*, 11225–11236.
44. Goldstein, H. *Classical Mechanics*; Addison-Wesley: Reading, MA, 1980.
45. Hockney, R. W. *Meth. Comput. Phys.* 1970, *9*, 136–211.
46. Verlet, L. *Phys. Rev.* 1967, *159*, 98–103.
47. Beeman, D. *J. Comput. Phys.* 1976, *20*, 130–139.
48. Berendsen, H. J. C.; van Gunsteren, W. F. Practical Algorithms for Dynamic Simulations. In *Molecular-Dynamics Simulation of Statistical-Mechanical Systems*; Ciccotti, G., Hoover, W. G., Eds.; North-Holland: Amsterdam, the Netherlands, 1986, pp 43–65.
49. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* 1977, *23*, 327–341.
50. van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* 1977, *34*, 1311–1327.
51. van Gunsteren, W. F.; Karplus, M. *Nature* 1981, *293*, 677–678.
52. Ehrenfest, P.; Ehrenfest, T. *Enzyklopädie der Mathematischen Wissenschaften IV*; Teubner: Leipzig, 1912; Vol. 2, pp 3–90.
53. Hünenberger, P. H. *Adv. Polym. Sci.* 2004, *173*, 105–149.
54. Owicki, J. C.; Scheraga, H. A. *J. Am. Chem. Soc.* 1977, *99*, 7403–7412.
55. Tembe, B. L.; McCammon, J. A. *Comput. Chem.* 1984, *8*, 281–283.
56. De Donder, T. *L'Affinité*; Gauther-Villars: Paris, 1927.
57. Mark, A. E.; van Gunsteren, W. F.; Berendsen, H. J. C. *J. Chem. Phys.* 1991, *94*, 3808–3816.
58. Resat, H.; Mezei, M. *J. Chem. Phys.* 1993, *99*, 6052–6061.
59. Pearlman, D. A. *J. Phys. Chem.* 1994, *98*, 1487–1493.
60. Kong, X.; Brooks, C. L., III. *J. Chem. Phys.* 1996, *105*, 2414–2423.
61. Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* 1974, *28*, 578–581.
62. Bennett, C. H. *J. Comput. Phys.* 1976, *22*, 245–268.
63. Tidor, B. *J. Phys. Chem.* 1993, *97*, 1069–1073.
64. Pitera, J. W.; Kollman, P. *J. Am. Chem. Soc.* 1998, *120*, 7557–7567.
65. Jarzynski, C. *Phys. Rev. Lett.* 1997, *78*, 2690–2693.
66. Jarzynski, C. *Phys. Rev. E* 1997, *56*, 5018–5035.
67. Hendrix, D. A.; Jarzynski, C. *J. Chem. Phys.* 2001, *114*, 5974–5981.
68. Hummer, G. *J. Chem. Phys.* 2001, *114*, 7330–7337.
69. Liu, H. Y.; Mark, A. E.; van Gunsteren, W. F. *J. Phys. Chem.* 1996, *100*, 9485–9494.
70. Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* 1994, *222*, 529–539.
71. Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* 1994, *100*, 9025–9031.
72. Oostenbrink, C.; van Gunsteren, W. F. *Proteins* 2004, *54*, 237–246.
73. Hummer, G.; Szabo, A. *J. Chem. Phys.* 1996, *105*, 2004–2010.
74. Bürgi, R.; Läng, F.; van Gunsteren, W. F. *Mol. Sim.* 2001, *27*, 215–236.
75. Åqvist, J.; Medina, C.; Samuelsson, J. E. *Protein Eng.* 1994, *7*, 385–391.
76. Åqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. *Acc. Chem. Res.* 2002, *35*, 358–365.
77. Hansson, T.; Marelius, J.; Åqvist, J. *J. Comput.-Aided Mol. Des.* 1998, *12*, 27–35.
78. Duffy, D. E.; Jorgensen, W. L. *J. Am. Chem. Soc.* 2000, *122*, 2878–2888.
79. Åqvist, J.; Hansson, T. *J. Phys. Chem.* 1996, *100*, 9512–9521.
80. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T. S.; Duan, Y.; Wang, W. et al. *Acc. Chem. Res.* 2000, *33*, 889–897.
81. Oprea, T. I.; Marshall, G. R. *Perspect. Drug Disc.* 1998, *9*, 35–61.
82. Böhm, H. J. *J. Comput.-Aided Mol. Des.* 1992, *6*, 61–78.
83. Lewis, R. A.; Leach, A. R. *J. Comput.-Aided Mol. Des.* 1994, *8*, 467–475.
84. Kuntz, I. D.; Meng, E. C.; Shoichet, B. K. *Acc. Chem. Res.* 1994, *27*, 117–123.
85. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* 1998, *19*, 1639–1662.
86. Sotriffer, C. A.; Flader, W.; Winger, R. H.; Rode, B. M.; Liedl, K. R.; Varge, J. M. *Methods* 2000, *20*, 280–291.
87. Marelius, J.; Ljungberg, K. B.; Åqvist, J. *Eur. J. Pharm. Sci.* 2001, *14*, 87–95.
88. Rami Reddy, M.; Viswanadhan, V. N.; Weinstein, J. N. *Proc. Natl. Acad. Sci. USA* 1991, *88*, 10287–10291.
89. Rao, B. G.; Tilton, R. F.; Singh, U. C. *J. Am. Chem. Soc.* 1992, *114*, 4447–4452.
90. Hulten, J.; Bonham, N. M.; Nillroth, U.; Hansson, T.; Zuccarello, G.; Bouzide, A.; Aqvist, J.; Classon, B.; Danielson, U. H.; Karlen, A. et al. *J. Med. Chem.* 1997, *40*, 885–897.
91. Marelius, J.; Hansson, T.; Åqvist, J. *Int. J. Quantum Chem.* 1998, *69*, 77–88.
92. McCarrick, M. A.; Kollman, P. A. *J. Comput.-Aided Mol. Design* 1999, *13*, 109–121.
93. Eriksson, M. A. L.; Pitera, J.; Kollman, P. A. *J. Med. Chem.* 1999, *42*, 868–881.
94. Lee, T. S.; Kollman, P. A. *J. Am. Chem. Soc.* 2000, *122*, 4385–4393.
95. Okimoto, N.; Tsukui, T.; Kitayama, K.; Hata, M.; Hoshino, T.; Tsuda, M. *J. Am. Chem. Soc.* 2000, *122*, 5613–5622.
96. Daura, X.; Haaksma, E.; van Gunsteren, W. F. *J. Comput.-Aided Mol. Des.* 2000, *14*, 507–529.
97. Pak, Y.; Enyedy, I. J.; Varady, J.; Kung, J. W.; Lorenzo, P. S.; Blumberg, P. M.; Wang, S. *J. Med. Chem.* 2001, *44*, 1690–1701.
98. Graffner-Nordberg, M.; Kolmodin, K.; Åqvist, J.; Queener, S. F.; Hallberg, A. *J. Med. Chem.* 2001, *44*, 2391–2402.
99. Rami Reddy, M.; Erion, M. D. *J. Am. Chem. Soc.* 2001, *123*, 6246–6252.
100. Brandsdal, B. O.; Aqvist, J.; Smalas, A. O. *Protein Sci.* 2001, *10*, 1584–1595.
101. van Lipzig, M. M. H.; ter Laak, A. M.; Jongejan, A.; Vermeulen, N. P. E.; Wamelink, M.; Geerke, D.; Meerman, J. H. N. *J. Med. Chem.* 2004, *47*, 1018–1030.

102. Park, H.; Lee, S. *J. Comput.-Aided Mol. Des.* 2005, *19*, 17–31.
103. Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. *J. Med. Chem.* 2005, *48*, 4040–4048.
104. Parker, M. G.; Arbuckle, N.; Dauvois, S.; Danielian, P.; White, R. *Ann. NY Acad. Sci.* 1993, *684*, 119–126.
105. Crisp, T. M.; Clegg, E. D.; Cooper, R. L.; Wood, W. P.; Anderson, D. G.; Baetcke, K. P.; Hoffman, J. L.; Morrow, M. S.; Rodier, D. J.; Schaeffer, J. E. et al. *Environ. Health Perspect.* 1998, *106*, 11–56.
106. Soto, A. M.; Sonnenschein, C.; Chung, K. L.; Fernandez, M. F.; Olea, N.; Serrano, F. O. *Environ. Health Perspect.* 1995, *103*, 113–122.
107. Waller, C. L.; Oprea, T. I.; Chae, K.; Park, H. K.; Korach, K. S.; Laws, S. C.; Wiese, T. E.; Kelce, W. R.; Earl Gray, L. *Chem. Res. Toxicol.* 1996, *9*, 1240–1248.
108. Breinholt, V.; Larsen, J. C. *Chem. Res. Toxicol.* 1998, *11*, 622–629.
109. Garner, C. E.; Jefferson, W. N.; Burka, L. T.; Matthews, H. B.; Newbold, R. R. *Toxicol. Appl. Pharmacol.* 1999, *154*, 188–197.
110. Blair, R. M.; Fang, H.; Branham, W. S.; Hass, B. S.; Dial, S. L.; Moland, C. L.; Tong, W. D.; Shi, L. M.; Perkins, R.; Sheehan, D. M. *Toxicol. Sci.* 2000, *54*, 138–153.
111. Anstead, G. M.; Carlson, K. E.; Katzenellenbogen, J. A. *Steroids* 1997, *62*, 268–303.
112. Charles, G. D.; Bartels, M. J.; Zacharewski, T. R.; Gollapudi, B. B.; Freshour, N. L.; Carney, E. W. *Toxicol. Sci.* 2000, *55*, 320–326.
113. van Lipzig, M. M. H.; Vermeulen, N. P. E.; Gusinu, R.; Legler, J.; Frank, H.; Seidel, A.; Meerman, J. H. N. *Environ. Toxicol. Pharmacol.* 2005, *19*, 41–55.
114. Adlercreutz, H.; Mazur, W. *Ann. Med.* 1997, *29*, 95–120.
115. Bergman, A.; Klassonwehler, E.; Kuroki, H. *Environ. Health Perspect.* 1994, *102*, 464–469.
116. Jansson, B.; Andersson, R.; Asplund, L.; Litzen, K.; Nylund, K.; Sellstrom, U.; Uvemo, U. B.; Wahlberg, C.; Wideqvist, U.; Odsjo, T. et al. *Environ. Toxicol. Chem.* 1993, *12*, 1163–1174.
117. Dewailly, E.; Nantel, A.; Weber, J. P.; Meyer, F. *Bull. Environ. Contam. Toxicol.* 1989, *43*, 641–646.
118. Sipes, I. G.; Schnellmann, R. G. Biotransformation of PCBS: Metabolic Pathways and Mechanisms. In *Polychlorinated Biphenyls (PCBs): Mammalian and Environmental Toxicology*; Safe, S., Ed.; Springer-Verlag: Heidelberg, Germany, 1987, pp 97–110.
119. Matthews, J.; Zacharewski, T. *Toxicol. Sci.* 2000, *53*, 326–339.
120. Brzozowski, A. M.; Pike, A. C. W.; Dauter, Z.; Hubbard, R. E.; Bonn, T.; Engstrom, O.; Ohman, L.; Greene, G. L.; Gustafsson, J. A.; Carlquist, M. *Nature* 1997, *389*, 753–758.
121. Danielan, P. S.; White, R.; Lees, J. A.; Parker, M. G. *EMBO J.* 1992, *11*, 1025–1033.
122. Oostenbrink, B. C.; Pitera, J. W.; Van Lipzig, M. M. H.; Meerman, J. H. N.; van Gunsteren, W. F. *J. Med. Chem.* 2000, *43*, 4594–4605.
123. Oostenbrink, C.; van Gunsteren, W. F. *Proc. Natl. Acad. Sci. USA* 2005, *102*, 6750–6754.
124. Shiau, A. K.; Barstad, D.; Loria, P. M.; Cheng, L.; Kushner, P. J.; Agard, D. A.; Greene, G. L. *Cell* 1998, *95*, 927–937.
125. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* 2000, *28*, 235–242.
126. Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* 1999, *103*, 3596–3607.
127. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, the Netherlands, 1981, pp 331–342.
128. Oosterkamp, A. J.; Villaverde Herraiz, M. T.; Irth, H.; Tjaden, U. R.; van der Greef, J. *Anal. Chem.* 1996, *68*, 1201–1206.
129. Buu-Hoï, N. P.; Ekert, B.; Royer, R. *J. Org. Chem.* 1954, *19*, 1548–1552.
130. Oostenbrink, C.; van Gunsteren, W. F. *Chem. Eur. J.* 2005, *11*, 4340–4348.
131. Yu, H. B.; Amann, M.; Hansson, T.; Köhler, J.; Wich, G.; van Gunsteren, W. F. *Carbohydr. Res.* 2004, *339*, 1697–1709.

Biographies

Chris Oostenbrink was born in 1977 in Amsterdam (The Netherlands). In 2000, he obtained MSc degrees in Chemistry and in Medicinal Chemistry at the Free University of Amsterdam. In 2004, he defended his PhD on free energy calculations in biomolecular simulation at the ETH in Zurich. From 2004 on he holds a position as assistant professor in the field of Computational Medicinal Chemistry and Toxicology at the Free University in Amsterdam.

Marola M H van Lipzig was born in Delft (The Netherlands). In 1998, she received her MSc degree in environmental studies from the Wageningen Agricultural University. Until 2003 she worked on her PhD thesis on environmental estrogens at the Free University in Amsterdam, followed by a postdoc in Computational Molecular Toxicology in 2004 and a postdoc in Medicinal Chemistry in 2005 at the same university.
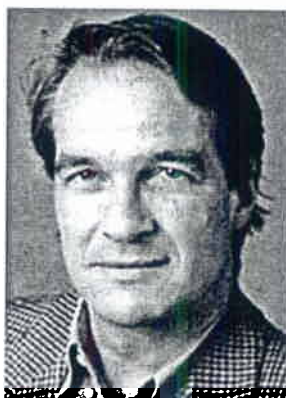
Wilfred F van Gunsteren was born in 1947 in Wassenaar (The Netherlands). In 1968, he gained a BSc in physics at the Free University of Amsterdam; in 1976 he was awarded a 'Meester' in Law, and in 1976 a PhD in nuclear physics. After postdoc research at the University of Groningen and at Harvard University he was, 1980–87, senior lecturer and, until August 1990, professor for physical chemistry at the University of Groningen. In 1990, he became professor of computer chemistry at the ETH Zurich. He is holder of a gold medal for research of the Royal Netherlands Chemical Society. His main interests center on the physical fundamentals of the structure and function of biomolecules.